DESCRIPTION

RHEUMATOID ARTHRITIS GENE AND METHOD FOR DIAGNOSING RHEUMATOID ARTHRITIS

Technical Field

The present invention relates to the disease gene of rheumatoid arthritis present in the human X chromosome and a method for diagnosing rheumatoid arthritis by detecting the presence of the disease gene or its expression product.

Background Art

Although aspects, particularly the pathological process, of arthritis and arthritis mutilans which cause rheumatoid arthritis, have been clarified through various investigations, because most autoimmune diseases associated with rheumatoid arthritis developed or worsen into the disease only when various causative factors coincide, the interaction itself of multiple factors must be clarified to understand the disease and to develop appropriate methods of treatment.

The number of patients with rheumatoid arthritis in the world is 1% or less (N. Engl. J. Med. 322: 1277-1289, 1990), but among sibilings of patients, over 8% develop the disease (Cell. 85: 311-318, 1996), which leads to the notion that some genetic factor may be involved. However, molecular genetic

procedures and genetic engineering processes used conventionally to discover the genetic factor of diseases may not be effective for autoimmune diseases. Such problem is caused by the fact that autoimmune diseases do not develop through mechanisms as simple as those of cancer, wherein abnormal growth of one mutated gene occurs. Further, although classical genetic procedures which search for genetic basis of a disease revealed that autoimmune diseases are caused by multiple genetic factors, it has not been successful in discovering its entrails or its body. Thus, almost nothing about the entity, or even the locus, of genes associated with rheumatoid arthritis has been known.

By performing linkage analysis using microsatellite markers on rheumatoid arthritis patients and their relatives, the present inventors identified three loci of rheumatoid arthritis genes (International Immunology 10(12): 1891-1895, 1998; Journal of Clinical Rheumatology 4(3): 156-158, 1998) and filed a patent application for the following disease genes (PCT/JP98/01665).

(1) A disease gene of rheumatoid arthritis located within ±1 centi Morgan vicinity of a DNA sequence on human chromosome 1 to which microsatellite marker(s) D1S214 and/or D1S253 hybridize(s).

(2) A disease gene of rheumatoid arthritis located within ±1 centi Morgan vicinity of a DNA sequence on human chromosome 8 to which microsatellite marker D8S556 hybridizes.

(3) A disease gene of rheumatoid arthritis located within ±1 centi Morgan vicinity of a DNA sequence on human chromosome X to which microsatellite marker(s) DXS1001, DXS1047, DXS1205, DXS1227 and/or DXS1232 hybridize(s).

The present inventors identified, as a result of further studies on each of the rheumatoid arthritis genes specified in the above-described previous application, the specific gene regarding the disease gene (3) described above and determined its molecular structure.

B Summary of the ~~Disclosure~~ Invention

Sub B1 In order to solve the above-described problems, the present invention provides a disease gene for rheumatoid arthritis, which is a mutant of protooncogene Dbl transcribing an mRNA that encodes the cDNA of which the sequence from the 2679th to 2952nd bases is shown in SEQ ID NO: 1, which disease gene transcribes an mRNA encoding the cDNA of which the region from the 20th to 274th bases in SEQ ID NO: 1 is substituted with the sequence of SEQ ID NO: 2.

Sub B2 The present invention also provides a cDNA of the above disease gene, a DNA fragment, which is a part of such cDNA, a protein expressed by the above disease gene, a peptide which is a part of such protein, and an antibody against such protein.

Further, the present invention provides a method for diagnosing rheumatoid arthritis comprising the detection of the

mRNA from the above disease gene or the above protein in a biological specimen.

The present invention further provides a method for the functionally complementing Dbl deficiency.

Description of the Preferred Embodiments

~~Best Modes for Carrying Out the Invention~~

Hereinafter, embodiments of the present invention having the above-described characteristics will be described.

The rheumatoid arthritis disease gene of the present invention (hereinafter referred to as "RA disease gene") is a variant sequence of known protooncogene Dbl gene (EMBO J. 7(8): 2463-2473, 1988; GenBank Accession No. X12556) which is isolated from human chromosome X by the method described in the after-mentioned Examples. In other words, this Dbl gene transcribes the mRNA encoding the cDNA for which the sequence of the 2679th to 2952nd bases is represented in SEQ ID NO: 1, while in the cDNA of the variant gene, the sequence of the 3' side of the 241st base in SEQ ID NO: 1 is linked to the downstream side of the 18th base to induce a frame shift in amino acid translation, causing the 19th to 274th base in SEQ ID NO: 1 to be substituted by the sequence shown in SEQ ID NO: 2. Fig. 1 shows the base sequence of the 2679th to 2952nd bases (same as SEQ ID NO: 1) of Dbl gene cDNA in a normal, the corresponding base sequence of RA disease gene, and the respective amino acid sequences (1 letter notation) encoded by these sequences.

In addition, generally, polymorphism of individual differences is often found for human genes. Thus, the RA disease gene of the present invention may include genes that code cDNAs obtained by the addition, deletion or substitution of one or more nucleotide in SEQ ID NO: 2. Likewise, the present invention also includes proteins with one or more amino acid added to, deleted from and/or substituted, produced by such change to the base.

The cDNAs of the present invention may easily be isolated by, for example, the method described in the after-mentioned Example. Further, the cDNAs of the present invention may be cloned from a cDNA library produced by a known method (Mol. Cell. Biol. 2:161-170, 1982; J. Gene 25: 263-269, 1983; Gene 150: 243-250, 1994) using poly(A)+RNA extracted from cells of a patient with rheumatoid arthritis. Such cloning may be performed by, for example, synthesizing oligonucleotides based on the sequence information provided by the present invention and screening by colony or plaque hybridization by a known method using the resultant oligonucleotides as probes. Also, oligonucleotides, which hybridize to both ends of the target cDNA fragment, may be synthesized, and using them as primers, the cDNA of the present invention may be produced by RT-PCR method from mRNAs isolated from cells of a patient with rheumatoid arthritis.

Sub B5 The DNA fragment of the present invention comprises a

portion of the aforesaid cDNA, and contains the base sequence shown in SEQ ID NO: 3. In other words, SEQ ID NO: 3 is the underlined sequence in Figure 1, and is a characteristic region, which is not present in normal Dbl gene or its cDNAs. Further, the DNA fragment includes both sense and antisense strands. These DNA fragments may be used as probes for genetic diagnosis.

The proteins of the present invention are expression products resulting from the RA disease genes of the present invention, and has the amino acid sequence shown in SEQ ID NO: 2 at its C-terminal. These proteins may be obtained by chemical peptide synthesis method based on the amino acid sequence provided by the present application, or by recombinant DNA technique using cDNAs provided by the present application. For example, when recombinant DNA technique is used to obtain the proteins, RNA may be prepared by *in vitro* transcription using a vector containing the cDNA of the present invention; using this RNA as a template, the proteins may be obtained by *in vitro* translation. Also, the coding region of the cDNA may be recombined into an appropriate expression vector by any known method, and the recombinant vector obtained may be used to transform *E. coli.*, *Bacillus subtilis*, yeast, animal cells or the like, whereby expression of the protein in bulk would be possible using these recombinant cells.

When *in vitro* translation is used to produce the proteins of the present invention, the coding region of the cDNA of the

present invention may be recombined into a vector with RNA polymerase promoter, and introduced into the *in vitro* translation system containing the RNA polymerase corresponding to the promoter, such as rabbit reticular erythrocyte lysate or wheat embryo extracts. T7, T3 and SP6 may be listed as examples of the RNA polymerase promoter. Examples of vectors, which contain any of these RNA polymerase promoters are pKA1, pCDM8, pT3/T7 18, pT7/3 19 and pBluescript II.

Furthermore, when the proteins of the present invention are expressed using microorganisms such as *E. coli.*, a recombinant expression vector may be prepared by incorporating the coding region of the cDNA of the present invention into an expression vector which contains replication origin replicable in microorganism, promoter, ribosome-binding site, cDNA cloning site, terminator and the like, which is then used to transform a host cell and incubating the transformed cell. In such cases, by adding initiation and termination codons before and after an arbitrary coding region, protein fragments, which contain the arbitrary region may be obtained. Alternatively, the protein may be obtained as a fusion protein with another protein. By cleaving the fusion protein using an appropriate protease, the target protein may also be isolated. Examples of the expression vector for *E. coli.* are pUC system, pBluescript II, pET expression system and pGEX expression system.

When expressing the protein of the present invention in

eucaryotic cells, the coding region of the inventive cDNA may be incorporated into an expression vector for eucaryotic cells that contains a promoter, a splicing region, a poly(A) addition site and the like, which may be introduced into eucaryotic cells. Such expression vectors may be pKA1, pCDM8, pSVK3, pMSG, pSVL, pBK-CMV, pBK-RSV, EBV vector, pRS and pYES2. Generally, mammal culture cells such as monkey kidney cell COS7 or Chinese hamster ovarian cell CHO, budding yeast, fission yeast, silkworm cells and *Xenopus laevis* o-site cells are used as eucaryotic cells, but in the present invention, they are not limited to these examples. To introduce the expression vector into eucaryotic cells, any known method such as electroporation, calcium phosphate method, liposome method, and DEAE dextran method may be used.

After the proteins are expressed in procaryotic or eucaryotic cells by the above-described methods, the protein of interest may be separated from the culture and purified by using combinations of known separation/purification methods. Examples are, treatment with degenerating agents such as urea or surfactant, ultrasonication, enzyme digestion, salt- or solvent-precipitation, dialysis, centrifugation, ultrafiltration, gel filtration, SDS-PAGE, isoelectric focusing method, ion-exchange chromatography, hydrophobic chromatography, affinity chromatography, reverse phase chromatography and the like.

Further, the protein of the present invention also encompasses fusion proteins of the present protein with other arbitrary protein.

The peptide of the present invention is a peptide fragment, which contains at least part (5 amino acid residues or more) of the amino acid sequence shown in SEQ ID NO: 2. Such peptide may be used as an antigen for preparing an antibody.

The antibody of the present invention may be obtained as a polyclonal or monoclonal antibody by any known method using the protein itself or a partial peptide thereof as antigen.

The method for diagnosing rheumatoid arthritis of the present invention may be performed, for example, by detecting the presence of characteristic mRNAs transcribed by RA disease gene in a biological specimen (body fluid, cell) obtained from a subject. Such mRNA may be detected by, for example, RT-PCR amplification of the mRNA containing the characteristic region (e.g., the underlined region in Figure 1), or by *in vitro* or *in situ* hybridization analysis using any characteristic sequence region of the mRNA for RA disease gene as a probe.

Furthermore, the method for diagnosing rheumatoid arthritis of the present invention may also be performed by detecting the presence of protein(s) expressed from RA disease gene in a biological specimen of a subject. Such detection may be performed by, for example, enzyme immunoassay or radioimmunoassay using the antibody of the present invention.

Further, the presence of such gene expression or protein may be detected by using any diagnosis kit; for example, hybridization analysis kit such as DNA chip and the like or immunoassay kit such as ELISA kit may be used.

The Dbl defect of the present invention may be complemented by, for example, protein or low molecular weight compounds.

Examples

Hereinafter, the RA disease gene of the present invention will be described in further detail through the following examples; however, the present invention is not limited to these examples.

<Example 1> Identification of the RA disease gene

For the gene analysis by affected sib-pair analysis method using microsatellite marker, DNAs were prepared from peripheral blood collected from a family of two rheumatoid arthritis patients and one normal, by the guanidine-thiocyanate method (The Japan Society of Blood Transfusion Report 40(2), 413). Further, 11 markers (DXS1047, DXS8072, DXS8041, DXS8094, DXS1192, DXS1205, DXS1227, DXS8106, DX8043, DX8028 and DXS1200) (Nature 360, 1996) were selected as microsatellite markers with heterozygosity higher than about 0.7, from the range of the candidate genetic loci previously disclosed by the present inventors (International Immunology 10(12): 1891-1895; Journal

of Clinical Rheumatology 4(3): 156-158, 1998), and fluorescence-labeled primers that could amplify each loci were synthesized at Perkin Elmer Inc. The sequences of the primer are disclosed in the above literature and are known. Each marker region was isolated by PCR under the following conditions. The reaction solution was prepared by mixing 5pmol of primer, approximately 0.5μg of template DNA, 1.5μg of Buffer II (Perkin Elmer Inc.), 1.0μl of 2mM dNTP Mix (Perkin Elmer Inc.), 0.12μl of Ampli Taq Gold enzyme (Perkin Elmer Inc.) and 0.9μl of 25mM $MgCl_2$ (Perkin Elmer Inc.), and adding sterilized water to obtain a total volume of 15μl. The reaction was performed in a thermal cycler (PTC-200) of MJ Research Inc. First, one cycle of enzyme activation at 95°C for 12 minute, 10 cycles of heat denaturation at 94°C for one minute, primer annealing at 47°C for one minute and extension at 72°C for 2 minutes were performed, after which 20 cycles of heat denaturation at 89°C for one minute, primer annealing at 47°C for one minute and extension at 72°C for 2 minutes were performed. Each of the resultant DNA fragments were analyzed in a DNA sequencer (Perkin Elmer Inc., Type AB1377) by subjecting to electrophoresis with size markers for Genescan (Perkin Elmer Inc.) of the manufacture's specification, and the DNA analysis was performed by using the attached softwares, Genescan and Genotyper. The data obtained were analyzed on Unix system using Mapmaker Sibs software (Am J Hum Genet, 57, 439-454, 1995), which is available to the public, for genetic linkage

analysis, and the maximum Lod value was calculated by single point analysis.

As a result, the maximum Lod was determined to be 2.03 for DXS984, which is located in the 0.1 centi Morgan vicinity of DXS1232, one of the candidate genetic loci disclosed by the present inventors (International Immunology 10(12): 1891-1895; Journal of Clinical Rheumatology 4(3): 156-158, 1998), showing significant correlation. By searching the international data base on the internet (Genemap98, http://www.ncbi.nlm.nih.gov/genemap98/), it was found that the physical location of DXS984 was 4259 cR10000(F) on the G3 Radiation hybrid map, and thus it was proved that the protooncogene Dbl was situated nearest to DXS984.

<Example 2> Analysis of Abnormal Dbl Gene

In order to compare the cDNAs between Dbl genes, cDNA was synthesized by reverse transcription using RT-PCR kit (Perkin Elmer Inc.) from the total RNA obtained from peripheral blood of RA disease patients collected using Isogen agent (Nippongene Co. Ltd.), and dissolved in 20μl of sterilized water. Furthermore, primers (SEQ ID NO: 4 and 5) were prepared using the Dbl cDNA sequence (Genbank Accession No. X12556) (Amersham Pharmacia), and part of the Dbl cDNA sequence was isolated by the PCR method. The composition of the reaction solution for PCR was: 10 pmol each of forward primer (SEQ ID NO: 4) and reverse primer (SEQ ID NO: 5), approximately 0.1μg of template DNA, 2.5μl

Sub B7 cont'd

of LA-PCR buffer (Takara Shuzo Co. Ltd.), 4.0µl of 2.5mM dNTP Mix, 0.25µl of LA Taq enzyme (Takara Shuzo Co. Ltd.) and 2.5µl of 25mM $MgCl_2$ mixed, after which sterilized water was added to obtain a total volume of 25µl. The reaction was performed in a thermal cycler (PTC-200) of MJ Research by repeating 35 cycles of the process of heat denaturation at 94°C for 30 seconds, primer annealing at 52°C for 30 seconds and extension at 72°C for 2 minutes. The PCR products were subjected to electrophoresis of conventional methods, in TAE buffer solution using 1% Agarose L (Nippongene Co. Ltd.) gel and DNA molecular weight markers (200bp ladder) by Promega Co., to confirm the amplified bands. As a result, it was found that the size of normal DNA was 660bp while the size of DNA chain from some patients were distinctly shorter (approximately 440bp).

Next, after each respective bands were cut out, the gels were melted at 65°C for 10 minutes, and the DNAs were purified by conventional phenol extraction methods and ethanol precipitation methods. Then, using 100ng of the resultant DNA as a template, cycle sequence reaction and purification were performed following the specifications of the manufacturer of BigDye terminator cycle sequence kit by Perkin Elmer Inc., and the sequence was determined by a Type AB1377 DNA sequencer of Perkin Elmer Inc. As a result, it was evident that in the above-described abnormally short DNA, as shown in Fig. 1, the 223bp from the number 2697 to number 2919 bases are deleted,

making it 437 bp. This result indicates that with the amino acid deletion encoded in the genetic information downstream of base number 2693, and by inducing frame shift, abnormal polypeptide chain short of 65 amino acids is produced.

Industrial Applicability

Sub B8 7

As described in detail above, the present invention provides a disease gene for rheumatoid arthritis occurring in human chromosome X. This invention enables the easy and reliable diagnosis of rheumatoid arthritis. Furthermore, this invention is useful for the development of novel treatment and therapeutic agents for rheumatoid arthritis.